Post-translational peptide splicing and T cell response

Michele Mishto[1,2,3,$] and Juliane Liepe[4,§]

[1] Institut für Biochemie, Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany

[2] Berlin Institute of Health, 10117 Berlin, Germany

[3] Centre for Inflammation Biology and Cancer Immunology (CIBCI) & Peter Gorer Department of Immunobiology, King's College London, SE1 1UL London, United Kingdom

[4] Max-Planck-Institute for Biophysical Chemistry, 37077 Göttingen, Germany

[$] To whom the correspondence should be addressed:

Michele Mishto, Institut für Biochemie, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. Tel: +49(0)30450528-288/071. Email: michele.mishto@kcl.ac.uk.

Juliane Liepe, Quantitative and Systems Biology group, Max-Planck-Institute for biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. Email: jliepe@mpibpc.mpg.de.

Running title: antigenic spliced peptides

Abstract

$CD8^+$ T cell-mediated specificity depends on the recognition of the MHC class I – epitope complexes at the cell surface. These epitopes are mainly produced upon the degradation of proteins by the proteasome, generating fragments of the original sequence. However, it is now clear that proteasome can produce a significant portion of epitopes by reshuffling the antigen sequence, thus expanding the potential antigenic repertoire. MHC class I-restricted spliced epitopes have been described in tumors and infections, suggesting an unpredicted relevance of these peculiar peptides. Here, we review current knowledge about proteasome-catalyzed peptide splicing, the emerging rules of this process, and the potential implications for our understanding and therapeutic use of $CD8^+$ T cells, as well as mechanisms generating other non-canonical antigenic epitopes targeted by T cell response.

The unpredicted frequency of proteasome-generated antigenic spliced peptides

$CD8^+$ T cell specificity is achieved through the recognition by the T cell receptor (TcR) of peptides presented in the cleft of **major histocompatibility complex class I** (MHC-I; see Glossary) molecules. The process of antigen presentation, from the original protein to the surface display of MHC-I molecule, has been the focus of intense studies. Since more than two decades, the proteasome is known to be the main protease responsible for the generation of the vast majority of peptides that are bound to the MHC-I cleft at the cell surface (**Box 1**). This macromolecular enzyme plays a critical role in protein homeostasis in all cells. However, each cell type can express different proteasome isoforms, which preferentially regulate specific pathways. For instance, the proteasome isoform expressed in immune cells or upon inflammation, *i.e.* the immunoproteasome, was initially thought to be specialized only in epitope generation, although later it became evident that this proteasome isoform is involved in several other inflammatory processes [1]. Another example is represented by the thymoproteasome, which impinges upon the mouse $CD8^+$ T cell repertoire by generating the immunopeptidome of the **cortical thymic epithelial cells** (cTECs), thereby regulating the thymocytes' positive selection, [2-5] (**Box 2**). Until 2004, all MHC-I-restricted epitopes known to the scientific community were directly derived from their corresponding antigen sequences, and, when being generated by proteasome, they were products of canonical peptide-bond hydrolysis. In 2004 however, two groups identified two epitopes, whose sequence was the fusion of two non-continuous segments of the tumor-associated **glycoprotein-100** (gp100) or the **fibroblast grow factor-5** (FGF-5) antigens carried out through **proteasome-catalyzed peptide splicing** (PCPS) [6-8]. In the following decade, the pioneering work of Vigneron, Van den Eynde and colleagues [9-11] led to the identification of only a few other proteasome-generated spliced epitopes (**Table 1**). Therefore, immunologists tagged PCPS as a rare event of marginal immunological relevance. The identification of this handful of spliced epitopes was possible because of the availability of **tumor-infiltrating $CD8^+$ T lymphocytes** (TILs) specific for the spliced epitopes. Conversely, other researchers developed reverse-immunology and bioinformatics approaches combined with mass spectrometry and database search strategies for the systematic identification of spliced peptides generated by proteasome. Novel methods were mandatory since the sequences of the spliced peptides were not included in any known proteome database, and the enormous number of theoretically possible combinations of peptide fragments that proteasome could ligate rendered the known proteomic approaches almost useless. These novel strategies allowed the

identification of several spliced peptides produced by proteasome during *in vitro* digestions of synthetic polypeptides [12-16], and the identification of the unforeseen dark side of the MHC-I immunopeptidome, *i.e.* thousands spliced peptides presented at the cells surface by MHC-I molecules of human B lymphoblastoid cell lines as well as primary fibroblasts [17].

The fact that almost one third of the antigenic peptides presented by different cells onto the MHC-I molecules are generated by PCPS was certainly surprising and might throw a spanner in the works that has important theoretical and therapeutic implications. Can this knowledge perturb our notion of self and non-self? How shall we reconsider the models of central and peripheral tolerance? Shall we reassess our most innovative strategies for epitope identification and validation used, for instance, for anti-cancer immunotherapies?

Here, we review first the mechanisms and regulation of PCPS, before focusing on its important ramifications.

Mechanisms and driving forces of PCPS

Proteasome-generated spliced peptides can be formed in different ways (**Fig. 1A, B**). The two splice-reactants are generated inside the proteasome chamber by cleavage of a single peptide sequence (*cis* PCPS), the excision of the sequence between the two splice-reactants (namely the intervening sequence) followed by the ligation of the splice-reactants in normal order, *i.e.* following the orientation from N- to C-terminus of the parental protein (normal *cis* PCPS), or in the reverse order (reverse *cis* PCPS). Furthermore, PCPS can also occur between two splice-reactants originating from two distinct proteins (*trans* PCPS). Although this occurs *in vitro* [8, 13, 15], its occurrence in cells is disputable [8]. It would however be difficult to reckon an immunological relevance of *trans* spliced peptides. Indeed, two specific antigens should repeatedly enter the proteasome in the same order and the putative formation of the *trans* spliced peptide should occur frequently enough to survive the reduction steps in the downstream MHC-I antigen presentation pathway and to trigger an efficient T cell response (**Box 1**). This could likely happen only in specific conditions (*e.g.* between viral antigens during viral infection) or in specific cellular compartments and/or macro-complexes, which would favor the reaction by molecular crowding of a limited variety of peptides. A similar speculation has been proposed to explain the production of hybrid epitopes able to trigger the $CD4^+$ T cell response in **type 1 diabetes mellitus** (T1DM) and generated by the fusion of fragments derived from distinct antigens by an unknown enzyme [18-20] (see below).

Despite the recent observation that PCPS can be carried out by a **condensation reaction** both *in vitro* (either in a cell-free system or in live cells) [16], the most studied and likely the most predominant PCPS mechanism is transpeptidation. According to the model originally proposed by Stroobant and colleagues [6] and later confirmed with diverse techniques [9, 13], the reaction starts with the break of the peptide-bond between the residue of the N-terminal splice reactant (*i.e.* the residue P_1) and the following intervening sequence, and the formation of an acyl-enzyme intermediate between the residue P_1 and the proteasome's catalytic Thr_1. During the canonical peptide hydrolysis, which produces all non-spliced peptides, the acyl-enzyme intermediate would react with a molecule of water leading to the release of the N-terminal (non-spliced) peptide fragment. In PCPS, on the contrary, the

nucleophilic attack to the acyl-enzyme intermediate is mediated by a C-terminal splice-reactant, thereby leading to the ligation between the residue P_1 of the N-terminal splice-reactant and the residue P_1' of the C-terminal splice-reactant (**Fig. 1C**). Since during the canonical peptide hydrolysis the substrate-binding sites of proteasome catalytic pockets are mandatory requirements for the efficient reaction catalyzed by the proteasome Thr_1, it has been hypothesized that also in PCPS specific substrate binding sites exist [10, 13, 21]. Whether these are the same binding sites used for the peptide hydrolysis still needs to be proven.

The existence of PCPS specific binding sites is, however, supported by preliminary evidences that PCPS is a finely tuned mechanism rather than a random process. In fact, the formation of given spliced peptides is more efficient when the splice-reactants have specific sequence characteristics [15] and the substrate's amino acid residues frequently used by proteasome for peptide hydrolysis are not those most frequently used for PCPS [13]. For one specific spliced epitope, the minimal length of the splice-reactant was shown to be 3 residues [10], although for other spliced peptides a single residue could be ligated in a transpeptidation reaction [22]. Furthermore, it has been shown for a specific spliced epitope that the longer is the intervening sequence, the less efficiently the spliced epitope is generated [8]. In the much larger pool of spliced peptides of the MHC-I immunopeptidomes studied so far, however, we do not see a correlation between the intervening sequence length and the antigenic spliced peptide frequency, although the method applied did not allow the identification of spliced peptides with intervening sequences longer than 20 amino acid residues [17].

Regarding the factors ruling PCPS, standard proteasome and immunoproteasome differ in the generation of specific spliced peptides [12, 13, 16, 23, 24]. This is not surprising since they have distinctive dynamics of the different aspects of their activities [25] (**Box 2**), although overall qualitative and/or quantitative differences of the PCPS carried out by the two isoforms have not been demonstrated so far [13]. How proteasome regulators such as 19S and PA28$\alpha\beta$ affect PCPS is still unknown.

Not just few molecules of each proteasome-generated spliced peptide

For immunologists, PCPS represents a spanner in the works not because of its existence, but because of its high frequency. Although the handful of spliced epitopes discovered in the last decade was an excellent pioneering work, it had no general consequences in translational medicine. On the contrary, the unexpected large variety and the abundance of the self spliced peptides eluted from the MHC-I molecules and identified by mass spectrometry - around one third of the immunopeptidome variety and one fourth of its abundance [17] – have some practical and theoretical implications. Regarding the practical implications, for instance, it will be now harder to annotate peptide sequences derived from MHC-I immunopeptidomes without considering spliced peptides. The latter would lead to the wrong assignment of 5-10% sequences, depending on the mass spectrometry method used [17]. Considering that the MHC-I immunopeptidome analysis is emerging as an extremely helpful tool for the identification of epitope targets for anti-cancer immunotherapy [26-28], the implications of PCPS in translational research are clear.

Regarding the theoretical implications, we have to consider that in immunobiology, quantity matters. Quantitative alterations of epitope production dynamics by the proteasome can lead to the complete abrogation of the $CD8^+$ T cell response in mouse models [29-32]. Therefore, determining the quantity of antigenic spliced peptides exposed to $CD8^+$ T cells by MHC-I molecules is a key element. Our quantification of the spliced and non-spliced peptides identified in three MHC-I immunopeptidomes was carried out by a label-free strategy based on the intensity of the mass spectrometry ion peak area [17]. Although this method is not directly suitable for quantitative comparison of single peptides [13, 33], it has become well accepted when analyzing large proteomics datasets to compare two groups of peptides making use of the law of large numbers [33-35]. Other quantitative methods [36] could be used to corroborate our recent results on three MHC-I immunopeptidomes, where spliced peptides showed an abundance distribution comparable to non-spliced peptides, although significantly lower [17]. A comparable abundance of spliced epitopes *vs* non-spliced epitopes at the cell surface has also been suggested by assays carried out with $CD8^+$ T cell clones specific for either spliced or non-spliced epitopes derived from melanoma-associated antigens [16]. In our experience, the latter strategy is more sensitive than modern mass spectrometry approaches - despite their sensational progresses in terms of accuracy and sensitivity obtained in the last years [37] – and could provide direct indications about the spliced peptides' immunogenicity. Nonetheless, the isolation of specific $CD8^+$ T cell clones is laborious and thus impractical for a system-wide analysis.

An alternative strategy could be the investigation of PCPS *in vitro* using purified proteasome and synthetic polypeptide substrates, which would in principle allow the quantification of a large number of spliced and non-spliced peptide products. Absolute quantification of the peptide products can be done through different mass spectrometry-based strategies [13, 37] and a correlation of quantitative analyses of proteasome-mediated digestions in cell-free systems with data obtained in cell culture and *in vivo* has been extensively reported [16, 23, 24, 31, 38-43]. The few *in vitro* studies on PCPS conducted so far were carried out with less informative mass spectrometry methods [13, 15] and we reckon that the use of more sensitive and accurate mass spectrometers will significantly enlarge our horizon about the frequency, relevance and characteristics of PCPS.

Proteasome-generated spliced epitopes derived from tumor-associated antigens

All six spliced epitopes validated so far in a human system derive from tumor-associated antigens (**Table 1**). For all of these spliced epitopes, human TILs or specific $CD8^+$ T cell clones have been isolated. Furthermore, a natural response of memory T cells derived from the peripheral blood of melanoma patients against autologous cells presenting spliced epitopes derived from the well-known melanoma-associated antigen gp100 – namely $gp100_{47\text{-}52/40\text{-}42}$ or $gp100_{195\text{-}202/192}$ – has been described [16].

In addition, the adoptive transfer of a TIL clone specific for the spliced epitope $tyrosinase_{368\text{-}373/336\text{-}340}$ into the autologous melanoma patient was followed by tumor regression [44]. In nonobese diabetic/severe combined immunodeficient mice (NOD/SCID), a $CD8^+$ T cell clone specific for $SP110_{296\text{-}301/286\text{-}289}$ inhibited engraftment of human acute myelogenous leukemia cells [45].

Although these studies all indicate that spliced epitopes could be useful targets for anti-cancer immunotherapies, the dimension of the MHC-I spliced immunopeptidome of tumor cells still needs to be determined. This is a critical aspect, which might have a significant impact on the most innovative anti-cancer immunotherapies such as **adoptive T cell therapies** (ATTs). The latter can use TILs, or autologous $CD8^+$ T cells expressing specific TCRs, as effectors for the recognition and elimination of cancer cells [46]. A significant determinant of the efficacy of ATTs resides in the identification of the "best" tumor antigens. Ideal targets are epitopes carrying driver somatic mutations, both because they represent tumor-specific neoantigens and because they are not easily lost by the tumor, since those mutations are involved in the process of malignant transformation [47]. The frequency of these tumor-specific mutated epitopes is, however, limited by the mutation frequency and by the fact that epitopes need to have not only specific motifs for being presented onto MHC-I molecules, but also for passing all steps of the antigen presentation pathway (**Box 1**). As a consequence, the majority of the TILs used in clinical trials target tumor-associated rather than tumor-specific antigens, which could in consequence trigger undesirable autoimmune reactions [48].

In these therapeutic frameworks, the inherent ability of spliced epitopes to extend the antigenic landscape and to have potentially extremely large sequence variety could be useful to target tumor-specific driver mutations that would not be efficiently exposed at the cell surface through non-spliced peptides because of incompatibility of the mutated protein sequence with the most common MHC-I variants. For instance, the HLA-A*02:01 complex, which is the predominant MHC-I variant in the Caucasian population, cannot efficiently present any of the non-spliced peptides that carry some of the most common cancer mutations, such as **BRAF** V600E or **KRAS** G12D. The latter, on the contrary, can be efficiently presented by a non-spliced epitope on the HLA-C*08:02 complex, which can be targeted by ATT thereby leading to a complete remission of metastatic colorectal cancer [49]. ATT strategies could also target a pool of spliced and non-spliced epitopes, which carry the same driver mutation and are presented onto the different MHC-I variants of the patient. This approach would prevent immune-escaping phenomena like the selective deletion of specific MHC-I-expressing genes [49].

Impact of PCPS during infections and implications for autoimmunity

$CD8^+$ T cell responses are often mandatory for the clearance of viruses and some other pathogens. Furthermore, vaccination trials for Dengue virus, Ebola virus or HIV showed how $CD8^+$ T cell activation can play a key role in vaccine development [50-52]. The evidences of the immunogenicity of pathogen-derived spliced epitopes during infection are preliminary, and limited so far to the intracellular pathogen *Listeria monocytogenes* [24, 53]. Although the immunological relevance of PCPS still needs to be demonstrated on a large scale, some evolutionary advantages of the human immune system to target pathogen-derived spliced epitopes can be easily estimated. For instance, PCPS significantly enlarges the number of self antigens represented onto the MHC-I molecules at the surface of human B lymphoblastoid cell lines or primary fibroblasts [17]. Similarly, during *Listeria monocytogenes* infection, the **phosphatidylcholine-preferring phospholipase C** (PlcB) antigen elicits a $CD8^+$ T cell response in the C57/BL6 mice only through the presentation of PlcB-derived spliced epitopes [54]. The potentially very large variety of the spliced epitopes could also be a useful strategy of the immune system to counter-attack mutation-based escape mechanisms of viruses.

Indeed, although a mutation can wipe out a single epitope [39], it is less likely to hamper the productions of an entire pool of spliced and non-spliced epitopes derived from that portion of the antigen.

However, it is important to consider that the same potentially very large variety of the spliced epitopes could influence autoimmune disease etiology. Calis and colleagues [55] estimated that the complete sequence overlap between *in silico* predicted antigenic peptides derived from self (human) or from a thousand different viruses is 0.15%, which raises to 0.7% considering the TCR recognition degeneracy. Of course, at the time that study did not include spliced peptides. It is easy to predict that the overlap between the self and non-self immunopeptidome will dramatically increase by encompassing PCPS in the frame. Therefore, theoretically, and depending on how PCPS is carried out in the thymus during T cell selection and maturation, we might have two extreme scenarios: either a scenario where autoimmunity is epidemic (due to a high overlap between self and non-self) or one where many viral antigens are not recognized by peripheral $CD8^+$ T cells, since the corresponding TCRs were deleted because of self-recognition, thereby further increasing the so-called holes in the T-cell repertoire [55]. Understanding why this is not the case requires investigating the link between PCPS and antigen presentation in thymus and in periphery.

At the moment, we can only speculate that PCPS could be involved in the $CD8^+$ T cell-mediated immune response against self antigens in those diseases (*e.g.* myelin in multiple sclerosis or insulin in T1DM), where infections have been hypothesized to be involved in the autoimmune pathogenesis [56, 57].

Hybrid insulin peptides in T1DM

In multiple sclerosis and T1DM, both $CD4^+$ and $CD8^+$ T cells can target their response toward self epitopes presented onto either MHC-I or MHC-II molecules [20, 57-60]. In particular, in T1DM, MHC-II-restricted and post-translationally spliced epitopes have been recently described. Indeed, two independent groups identified hybrid epitopes – *i.e. trans* spliced epitopes according to our nomenclature - that are specifically recognized by $CD4^+$ T cells derived from NOD mice or pancreatic islets of patients affected T1DM [18-20]. These hybrid peptides derive from the ligation of few sequences of proinsulin to fragments of antigens immunologically relevant in T1DM and packed in the secretory vesicles [19, 20]. One $CD4^+$ T cell clone, grown directly from an islet of a T1DM patient, recognizes a hybrid peptide derived from the ligation of the insulin C peptide and the insulin A chain. This specific epitope could thus be either a *trans* or a *cis* spliced peptide, depending on whether it is generated after of before the processing of the proinsulin that lead to the formation of the mature insulin form [20].

Because of the targeted-sequence strategy applied in both studies, it is still unclear which is the general prevalence of antigenic spliced (both *cis* and *trans*) peptides bound to the MHC-II molecules, and whether the phenomenon is relevant only for T1DM. Furthermore, it is still unknown which enzyme catalyzes their production. Beside the proteasome, other proteases such as Trypsin, Lys-C and plant-derived Asparaginyl endoprotease are known to catalyze peptide splicing [61]. In particular, the human homologous of the latter protease is also involved in the generation of MHC-II-restricted

epitopes [62], and thus it could be a validate candidate for the production of hybrid insulin peptides as well as other spliced epitopes presented onto MHC-II complexes.

Concluding remarks

Bearing in mind the enormous theoretical number of potential combination of splice-reactants, we reckon that algorithms estimating the likelihood of a spliced peptide to be generated and to bind to the MHC-I molecules will soon be mandatory to investigate the therapeutic applications of PCPS (see Outstanding Questions). Although preliminary evidences that PCPS is not a random process are encouraging, the pioneering studies on PCPS should be validated on much larger spliced peptide pools than the few peptides used as preliminary models. Furthermore, the present MHC-I-peptide binding affinity prediction algorithms should be applied carefully to spliced peptides, because they have mainly been trained using non-spliced peptide databases, and therefore their use could lead to a large number of false negatives [17]. The prediction of immunologically relevant spliced epitopes will likely be aided by the analysis of the corresponding antigen characteristics, abundance, turnover and translation rate, as these parameters have been correlated with non-spliced immunopeptidomes and intracellular proteomes in recent brilliant studies [28, 33, 63-66].

Despite the large number of spliced peptides identified in MHC-I immunopeptidomes, and the current focus on PCPS, we shall not forget that spliced peptides are only one part of the "non-canonical" peptides that are presented onto MHC molecules. In addition to the recent identification of the MHC-II-restricted hybrid insulin epitopes, which has been here reviewed, there are studies suggesting that a sizeable number of MHC-I-restricted antigenic peptides can derive from alternative open reading frames [67]. For instance, in T1DM patients, Kracht and colleagues described cytotoxic $CD8^+$ T cells capable of killing human β cells by targeting an epitope generated by the usage of an alternative open reading frame in the insulin gene [68]. Furthermore, MHC-I-restricted epitopes can originate from non-coding sequences of the genome [67, 69], and, in principle, also from intergenically spliced chimeric RNAs [70].

Understanding the extension and the immunological relevance of the non-canonical "dark side" of the immunopeptidome will be a great challenge with potentially strong translational implications.

Acknowledgment

This work was supported in part by Berlin Institute of Health (BIH, CRG1-TP1).

References

1. Groettrup, M. et al. (2010) Proteasomes in immune cells: more than peptide producers? Nat Rev Immunol 10 (1), 73-8.
2. Xing, Y. et al. (2013) Thymoproteasome subunit-beta5T generates peptide-MHC complexes specialized for positive selection. Proc Natl Acad Sci U S A 110 (17), 6979-84.
3. Murata, S. et al. (2007) Regulation of CD8+ T cell development by thymus-specific proteasomes. Science 316 (5829), 1349-53.

4. Kincaid, E.Z. et al. (2016) Specialized proteasome subunits have an essential role in the thymic selection of CD8(+) T cells. Nat Immunol 17 (8), 938-45.
5. Nitta, T. et al. (2010) Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. Immunity 32 (1), 29-40.
6. Vigneron, N. et al. (2004) An antigenic peptide produced by peptide splicing in the proteasome. Science 304 (5670), 587-90.
7. Hanada, K. et al. (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. Nature 427 (6971), 252-6.
8. Dalet, A. et al. (2010) Splicing of distant Peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. J Immunol 184 (6), 3016-24.
9. Warren, E.H. et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. Science 313 (5792), 1444-7.
10. Michaux, A. et al. (2014) A spliced antigenic peptide comprising a single spliced amino acid is produced in the proteasome by reverse splicing of a longer peptide fragment followed by trimming. J Immunol 192 (4), 1962-71.
11. Dalet, A. et al. (2011) An antigenic peptide produced by reverse splicing and double asparagine deamidation. Proc Natl Acad Sci U S A.
12. Textoris-Taube, K. et al. (2015) The T210M Substitution in the HLA-a*02:01 gp100 Epitope Strongly Affects Overall Proteasomal Cleavage Site Usage and Antigen Processing. J Biol Chem 290 (51), 30417-28.
13. Mishto, M. et al. (2012) Driving Forces of Proteasome-catalyzed Peptide Splicing in Yeast and Humans. Mol Cell Proteomics 11 (10), 1008-23.
14. Liepe, J. et al. (2010) The 20S Proteasome Splicing Activity Discovered by SpliceMet. PLOS Computational Biology 6 (6), e1000830.
15. Berkers, C.R. et al. (2015) Definition of Proteasomal Peptide Splicing Rules for High-Efficiency Spliced Peptide Presentation by MHC Class I Molecules. J Immunol 195 (9), 4085-95.
16. Ebstein, F. et al. (2016) Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. Sci Rep 6, 24032.
17. Liepe, J. et al. (2016) A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science 354 (6310), 354-358.
18. Wiles, T.A. et al. (2017) An insulin-IAPP hybrid peptide is an endogenous antigen for CD4 T cells in the non-obese diabetic mouse. J Autoimmun 78, 11-18.
19. Delong, T. et al. (2016) Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. Science 351 (6274), 711-4.
20. Babon, J.A. et al. (2016) Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. Nat Med 22 (12), 1482-1487.
21. Borissenko, L. and Groll, M. (2007) Diversity of proteasomal missions: fine tuning of the immune response. Biol Chem 388 (9), 947-55.
22. Berkers, C.R. et al. (2015) Peptide Splicing in the Proteasome Creates a Novel Type of Antigen with an Isopeptide Linkage. J Immunol 195 (9), 4075-84.
23. Dalet, A. et al. (2011) Differences in the production of spliced antigenic peptides by the standard proteasome and the immunoproteasome. Eur J Immunol 41 (1), 39-46.
24. Platteel, A.C. et al. (2016) CD8 T cells of Listeria monocytogenes-infected mice recognize both linear and spliced proteasome products. Eur J Immunol.
25. Liepe, J. et al. (2015) Quantitative time-resolved analysis reveals intricate, differential regulation of standard- and immuno-proteasomes. Elife 4.
26. Gubin, M.M. et al. (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature 515 (7528), 577-81.

27. Berlin, C. et al. (2015) Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy. Leukemia 29 (3), 647-59.
28. Bassani-Sternberg, M. et al. (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 7, 13404.
29. Mishto, M. et al. (2014) Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation. Eur J Immunol 44 (12), 3508-21.
30. Kincaid, E.Z. et al. (2011) Mice completely lacking immunoproteasomes show major changes in antigen presentation. Nat Immunol.
31. Deol, P. et al. (2007) Rates of processing determine the immunogenicity of immunoproteasome-generated epitopes. J Immunol 178 (12), 7557-62.
32. Zanker, D. et al. (2013) Mixed Proteasomes Function To Increase Viral Peptide Diversity and Broaden Antiviral CD8+ T Cell Responses. J Immunol 191 (1), 52-9.
33. Bassani-Sternberg, M. et al. (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 14 (3), 658-73.
34. de Graaf, E.L. et al. (2014) Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. Mol Cell Proteomics 13 (9), 2426-34.
35. Cox, J. et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13 (9), 2513-26.
36. Caron, E. et al. (2015) An open-source computational and data resource to analyze digital maps of immunopeptidomes. Elife 4.
37. Lossl, P. et al. (2016) The diverse and expanding role of mass spectrometry in structural and molecular biology. EMBO J 35 (24), 2634-2657.
38. Platteel, A.C.M. et al. (2016) Strategies to enhance immunogenicity of cDNA vaccine encoded antigens by modulating antigen processing. Vaccine.
39. Tenzer, S. et al. (2009) Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. Nat Immunol 10 (6), 636-46.
40. Guillaume, B. et al. (2012) Analysis of the processing of seven human tumor antigens by intermediate proteasomes. J Immunol 189 (7), 3538-47.
41. Chapiro, J. et al. (2006) Destructive cleavage of antigenic peptides either by the immunoproteasome or by the standard proteasome results in differential antigen presentation. J Immunol 176 (2), 1053-61.
42. Belogurov, A., Jr. et al. (2015) Ubiquitin-independent proteosomal degradation of myelin basic protein contributes to development of neurodegenerative autoimmunity. Faseb J 29 (5), 1901-13.
43. Schwarz, K. et al. (2000) Overexpression of the proteasome subunits LMP2, LMP7, and MECL-1, but not PA28 alpha/beta, enhances the presentation of an immunodominant lymphocytic choriomeningitis virus T cell epitope. J Immunol 165 (2), 768-78.
44. Robbins, P.F. et al. (1994) Recognition of tyrosinase by tumor-infiltrating lymphocytes from a patient responding to immunotherapy. Cancer Res 54 (12), 3124-6.
45. Bonnet, D. et al. (1999) CD8(+) minor histocompatibility antigen-specific cytotoxic T lymphocyte clones eliminate human acute myeloid leukemia stem cells. Proc Natl Acad Sci U S A 96 (15), 8639-44.
46. Rosenberg, S.A. and Restifo, N.P. (2015) Adoptive cell transfer as personalized immunotherapy for human cancer. Science 348 (6230), 62-8.
47. Blankenstein, T. et al. (2015) Targeting cancer-specific mutations by T cell receptor gene therapy. Curr Opin Immunol 33, 112-9.
48. Coulie, P.G. et al. (2014) Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. Nat Rev Cancer 14 (2), 135-46.

49. Tran, E. et al. (2016) T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. N Engl J Med 375 (23), 2255-2262.
50. Screaton, G. et al. (2015) New insights into the immunopathology and control of dengue virus infection. Nat Rev Immunol 15 (12), 745-59.
51. Haynes, B.F. (2015) New approaches to HIV vaccine development. Curr Opin Immunol 35, 39-47.
52. Ewer, K.J. et al. (2016) Viral vectors as vaccine platforms: from immunogenicity to impact. Curr Opin Immunol 41, 47-54.
53. Platteel, A.C. et al. (In revision) Multi-level strategy identifies spliced epitopes targeted by CD8+ T cells in bacterial infection. Cell Rep.
54. Platteel, A.C. et al. (In revision) Multi-level strategy identifies spliced epitopes targeted by CD8+ T cells during bacterial infection. Cell Rep.
55. Calis, J.J. et al. (2012) Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. PLoS Comput Biol 8 (3), e1002412.
56. Handel, A.E. et al. (2010) Environmental factors and their timing in adult-onset multiple sclerosis. Nat Rev Neurol 6 (3), 156-66.
57. Bulek, A.M. et al. (2012) Structural basis for the killing of human beta cells by CD8(+) T cells in type 1 diabetes. Nat Immunol 13 (3), 283-9.
58. Unger, W.W. et al. (2011) Discovery of low-affinity preproinsulin epitopes and detection of autoreactive CD8 T-cells using combinatorial MHC multimers. J Autoimmun 37 (3), 151-9.
59. Skowera, A. et al. (2008) CTLs are targeted to kill beta cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. J Clin Invest 118 (10), 3390-402.
60. Handel, A.E. et al. (2009) Type 1 diabetes mellitus and multiple sclerosis: common etiological features. Nat Rev Endocrinol 5 (12), 655-64.
61. Berkers, C.R. et al. (2009) Transpeptidation and reverse proteolysis and their consequences for immunity. Int J Biochem Cell Biol 41 (1), 66-71.
62. Manoury, B. et al. (1998) An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. Nature 396 (6712), 695-9.
63. Pearson, H. et al. (2016) MHC class I-associated peptides derive from selective regions of the human genome. J Clin Invest 126 (12), 4690-4701.
64. McShane, E. et al. (2016) Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. Cell 167 (3), 803-815 e21.
65. Bourdetsky, D. et al. (2014) The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A 111 (16), E1591-9.
66. Boisvert, F.M. et al. (2012) A quantitative spatial proteomics analysis of proteome turnover in human cells. Mol Cell Proteomics 11 (3), M111 011429.
67. Laumont, C.M. et al. (2016) Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. Nat Commun 7, 10238.
68. Kracht, M.J. et al. (2017) Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. Nat Med 23 (4), 501-507.
69. Apcher, S. et al. (2013) Translation of pre-spliced RNAs in the nuclear compartment generates peptides for the MHC class I pathway. Proc Natl Acad Sci U S A 110 (44), 17951-6.
70. Jia, Y. et al. (2016) Intergenically Spliced Chimeric RNAs in Cancer. Trends Cancer 2 (9), 475-484.
71. Granados, D.P. et al. (2015) The nature of self for T cells-a systems-level perspective. Curr Opin Immunol 34, 1-8.
72. Rock, K.L. et al. (2014) Re-examining class-I presentation and the DRiP hypothesis. Trends Immunol 35 (4), 144-52.

73. Anton, L.C. and Yewdell, J.W. (2014) Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. J Leukoc Biol 95 (4), 551-62.
74. Weinzierl, A.O. et al. (2008) Features of TAP-independent MHC class I ligands revealed by quantitative mass spectrometry. Eur J Immunol 38 (6), 1503-10.
75. Wei, M.L. and Cresswell, P. (1992) HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. Nature 356 (6368), 443-6.
76. Rock, K.L. et al. (2016) Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 37 (11), 724-737.
77. Ben-Nissan, G. and Sharon, M. (2014) Regulating the 20S proteasome ubiquitin-independent degradation pathway. Biomolecules 4 (3), 862-84.
78. Ruschak, A.M. and Kay, L.E. (2012) Proteasome allostery as a population shift between interchanging conformers. Proc Natl Acad Sci U S A 109 (50), E3454-62.
79. Dick, T.P. et al. (1996) Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. Cell 86 (2), 253-62.
80. Vigneron, N. and Van den Eynde, B.J. (2014) Proteasome subtypes and regulators in the processing of antigenic peptides presented by class I molecules of the major histocompatibility complex. Biomolecules 4 (4), 994-1025.
81. Emmerich, N.P. et al. (2000) The human 26 S and 20 S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. J Biol Chem 275 (28), 21140-8.
82. Arciniega, M. et al. (2014) Differential global structural changes in the core particle of yeast and mouse proteasome induced by ligand binding. Proc Natl Acad Sci U S A 111 (26), 9479-84.
83. Guillaume, B. et al. (2010) Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules. Proc Natl Acad Sci U S A 107 (43), 18599-604.
84. Toes, R.E. et al. (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med 194 (1), 1-12.
85. Sasaki, K. et al. (2015) Thymoproteasomes produce unique peptide motifs for positive selection of CD8(+) T cells. Nat Commun 6, 7484.
86. Ebstein, F. et al. (2012) Emerging roles of immunoproteasomes beyond MHC class I antigen processing. Cell Mol Life Sci.
87. Xu, M.M. et al. (2016) The Role of Adaptive Immunity in the Efficacy of Targeted Cancer Therapies. Trends Immunol 37 (2), 141-53.
88. Prior, I.A. et al. (2012) A comprehensive survey of Ras mutations in cancer. Cancer Res 72 (10), 2457-67.

Figure 1. Proteasome-catalyzed peptide splicing. Spliced peptides can be formed by: (**A**) *cis* PCPS, *i.e.* when the two splice-reactants derive from the same polypeptide molecule. The ligation of the splice-reactants can occur in normal order, *i.e.* following the orientation from N- to C-terminus of the parental protein (normal *cis* PCPS), or in the reverse order (reverse *cis* PCPS); (**B**) *trans* PCPS, *i.e*. when the two splice-reactants originate from two distinct protein molecules or two distinct proteins [8, 13]. **C**) The most defined PCPS mechanism is the transpeptidation. According to this model, the proteasome's catalytic Thr_1 breaks the peptide bond of the residue (P1) of the protein - thereby forming an acyl-enzyme intermediate with the N-terminal splice-reactant, coupled to the release of the intervening sequence – and, instead of catalyzing the canonical peptide hydrolysis, it catalyzes the ligation between the P1 residue of the N-terminal splice-reactant with the residue P1´ of the C-terminal splice-reactant [6, 13].

Table 1. Spliced epitopes derived from tumor-associated antigens

epitope	**antigen**	**identification method**	**notes**	**Reference**
[RTK][QLYPEW]	$gp100_{40\text{-}42/47\text{-}52}$	$CD8^+$ T cell clone		[6]
[NTYAS][PRFK]	$FGF\text{-}5_{172\text{-}176/217\text{-}220}$	isolated TIL		[7]
[SLPRGT][STPK]	$SP110_{296\text{-}301/286\text{-}289}$	$CD8^+$ T cell clone	specific $CD8^+$ T cell clone inhibited engraftment of human acute myelogeneous leukemia cells in mouse model	[9]
[IYMDGT][ADFSF]	$tyrosinase_{368\text{-}373/336\text{-}340}$	isolated TIL	tumor regression in melanoma patient after ATT with the TIL	[11]
[RSYVPLAH][R]	$gp100_{195\text{-}202/192}$	isolated TIL	natural response of memory $CD8^+$ T cells from peripheral blood of melanoma patients	[10, 16]
[QLYPEW][RTK]	$gp100_{47\text{-}52/40\text{-}42}$	*in vitro* digestions	natural response of memory $CD8^+$ T cells from peripheral blood of melanoma patients	[16]

Text box 1

Proteasome-mediated MHC-I-restricted antigen presentation pathway.

The peptides presented at the cell surface on MHC-I molecules represent the immunopeptidome of the cell, which is specific to a given cell and it is influenced by several factors such as its MHC-I haplotype (each diploid human cell has six usually distinct MHC-I isoforms), and its intracellular proteome, *i.e.* the proteins expressed in the cell [71]. If an antigenic peptide bound to the MHC-I complex is recognized by TCRs of $CD8^+$ T cells, it is defined as epitope. The large majority of the MHC-I-restricted antigenic peptides are produced in the proteasome chamber by canonical peptide-bond hydrolysis or, as discussed here, by PCPS. These peptides can derive from rapidly degraded proteins (RDPs), which are supposed to include defective ribosomal products (DRiPs) and non-exponentially degraded proteins (NEDPs), although the quantitative analysis of the immunopeptidome's source is still a matter of debate [33, 64, 65, 71-73]. In the cytosol, the peptide fragments released from the proteasome are further degraded by aminopeptidases, thereby regenerating the cellular amino acids pool. Few of these peptides survive this step and are transported into the endoplasmic reticulum (ER) through a channel constituted by the transporters associated with antigen processing (TAPs). TAP-independent pathways, which are only partially proteasome-dependent, have also been described [74, 75]. In the ER, the peptides are bound to the peptide-loading complex, can be further trimmed by ER-resident aminopeptidases (ERAPs) and can finally bind the MHC-I complexes. The MHC-I-peptide complex undergoes modifications, and is transported through the Golgi to the cell surface. There, it is presented for a variable period of time, which depends in part on the affinity between the MHC-I complex and the peptide, as empty MHC-I molecules are rapidly recycled. At the cell surface, the MHC-I-peptide complex can be recognized by the TCR and CD8 of T cells, and, in presence of co-stimulatory factors (*e.g.* CD28, CD80, CD86), the binding induces the priming/activation of T cells [76].

Text box 2

Proteasome isoforms and functions

The 20S proteasome is the core of the ubiquitin-proteasome system (UPS), which is responsible for the degradation of the large majority of the cytosolic proteins. In eukaryotic cell, the complete inhibition of proteasome activity induces cell death through necrosis and apoptosis. Although there are evidences of the function of the 20S proteasome alone [77], this protease is often associated to regulatory particles such as 19S, PA28$\alpha\beta$, PA28γ, PA200, which affect proteasome conformation and activity [78-81]. Different 20S proteasome isoforms exist. The most common is the standard proteasome, whose catalytic activity is displayed by the Thr_1 of the β1, β2 and β5 subunits. The proteasome in which these subunits are replaced by the β1i/LMP2, β2i/MECL1 and β5i/LMP7 subunits is named immunoproteasome and it is present in immune cells and when cells are exposed to an inflammatory milieu. Between the standard proteasome and immunoproteasome, there are also intermediate-type 20S proteasomes, which carry a different combination of standard- and immuno-subunits [80]. Of note, cells generally contain a mixture of different 20S proteasome isoforms. In 2007,

a novel proteasome isoform has been identified, which is specific for the thymic cortex, *i.e.* the thymoproteasome, which carries the β1i, β2i and β5t subunits [3].

The effects of the changes in the catalytic subunit composition of standard proteasome and immunoproteasome have been studied by applying a large variety of methods, whereas for the thymoproteasome the research mainly focused so far on its role to determine the T cell repertoire in mice.

The modifications of the catalytic subunits alter, of course, the catalytic pocket and thus the cleavage preferences but also they induce subtle modifications in the proteasome conformation, which affect proteasome proteolytic dynamics with consequences on protein turnover and on which and/or in which amount peptide products are produced [25, 29, 40, 41, 82-85].

For an overview of proteasome isoform functions, we suggest a few outstanding reviews [1, 80, 86].

Glossary

adoptive T cell therapies (ATTs): is a highly personalized cancer therapy that involves administration of T cells with direct anti-cancer activity to the cancer-bearing host [46].

BRAF: an isoform of RAF. RAF proteins are intermediate to Ras and MAPK in the cellular proliferative pathway. The BRAF V600E is the most common oncogenic mutation of BRAF in cancer. The mutation results in constitutive activation of the BRAF kinase and promotes cell transformation [87].

condensation reaction: in this review's context, is a peptide splicing reaction catalyzed by proteasome, which does not need the initial peptide-bond cleavage (*i.e*. the step 1 in **Fig. 1C**).

cortical thymic epithelial cells (cTECs): antigen presenting cells involved in the positive selection of the thymocytes in the cortical portion of the thymus. In mouse, they express mainly thymoproteasome, which has β1i, β2i, β5t as catalytic subunits [3].

fibroblast grow factor-5 (FGF-5): a secreted protein that is member of a family of heparin-binding growth factors. It regulates cell proliferation and it is overexpressed by multiple adenocarcinomas, thereby representing a tumor-associated antigen and a potential immunotherapy target.

glycoprotein-100 (gp100): gp100, also known as melanocyte protein PMEL, is a tumor-associated antigen expressed by the majority of malignant melanomas. It has been shown to be immunogenic and elicits an antigen-specific $CD8^+$ T cell response. It is currently used in anti-melanoma immunotherapy trials.

KRAS: one of the Ras proteins, which are proto-oncogenes that are frequently mutated in human cancers. HRas, KRas, and NRas are GTPase that function as molecular switches in regulating pathways that are responsible for proliferation and cell survival. The KRAS G12D is a very frequent mutation in colon carcinomas and other cancers; it favors GTP binding and produces constitutive activation of Ras [88].

major histocompatibility complex class I (**MHC-I)**: is the complex that presents epitopes to the TcR of $CD8^+$ T cells the epitopes. The recognition of the MHC-I-epitope complex by the TcR of $CD8^+$ T cells is the pre-requisite for the priming and/or activation of the $CD8^+$ T cells (in presence of other co-factors). The MHC-I complex is constituted of a molecule of β2-microglobulin and the MHC-I heavy chain, and is expressed by almost all cells but erythrocytes.

phosphatidylcholine-preferring phospholipase C (PlcB): *Listeria monocytogenes* is a Gram-positive bacterium that primarily infects phagocytes and then mobilizes the host cell cytoskeleton to spread to neighboring cells. To enter the cytosol of infected cells, the bacteria secrete LLO, the phospholipases PlcA/phosphatidylinositol-specific phospholipase C (PI-PLC) and

PlcB/phosphatidylcholine-preferring phospholipase C (PC-PLC). Bacterial clearance during *L. monocytogenes* infection is mediated by $CD8^+$ T cells specific for the secreted bacterial proteins.

tumor-infiltrating $CD8^+$ T lymphocytes (TILs): are CD8+ T cells present in the tumor. TILs are one of the main source of T cells for ATTs. They can specifically recognize tumor-associated antigens and tumor-specific epitopes.

proteasome-catalyzed peptide splicing (PCPS): the activity of proteasomes whereby two non-contiguous fragments of a polypeptide substrate or a protein are ligated, thereby generating a spliced peptide with a sequence not present in the original protein.

type 1 diabetes mellitus (T1DM): an autoimmune disease mediated by T cells responding to self antigens (*e.g.* proinsulin and islet amyloid polypeptide) in the pancreatic β cells of islets of Langherans. Both $CD4^+$ and $CD8^+$ T cell responses play a central role in the β cell destruction, which leads to insulin deficiency and hyperglycemia.

Outstanding questions

Which biochemical characteristics of proteasome catalytic sites favor PCPS efficiency so much more than expected? How are they altered by proteasome regulators such as 19S or PA28$\alpha\beta$?

Do tumor cells carry out PCPS as efficiently as lymphoblastoid B cells and primary human fibroblasts?

Does cell infection affect proteasome dynamics and regulate PCPS efficiency? Can proteasome-generated spliced epitopes derived from pathogens be as immunodominant as non-spliced epitopes?

What is the frequency and abundance of proteasome-generated spliced epitopes at the cell surface of cTECs, mTECs and other professional antigen presenting cells involved in the positive/negative selection of thymocytes? How does the immune system cope with the larger antigen landscape generated by PCPS? Is the latter somehow related to the onset / development of any autoimmune response?

Is it worth investing time and resources to identify spliced epitopes and target them through immunotherapies?

Is PCPS relevant only for the MHC-I-restricted antigen presentation or does it have other functions either inside or outside the cell?

Which are the real dimension and immunological relevance of the non-canonical MHC-I and –II immunopeptidomes?

Trends Box

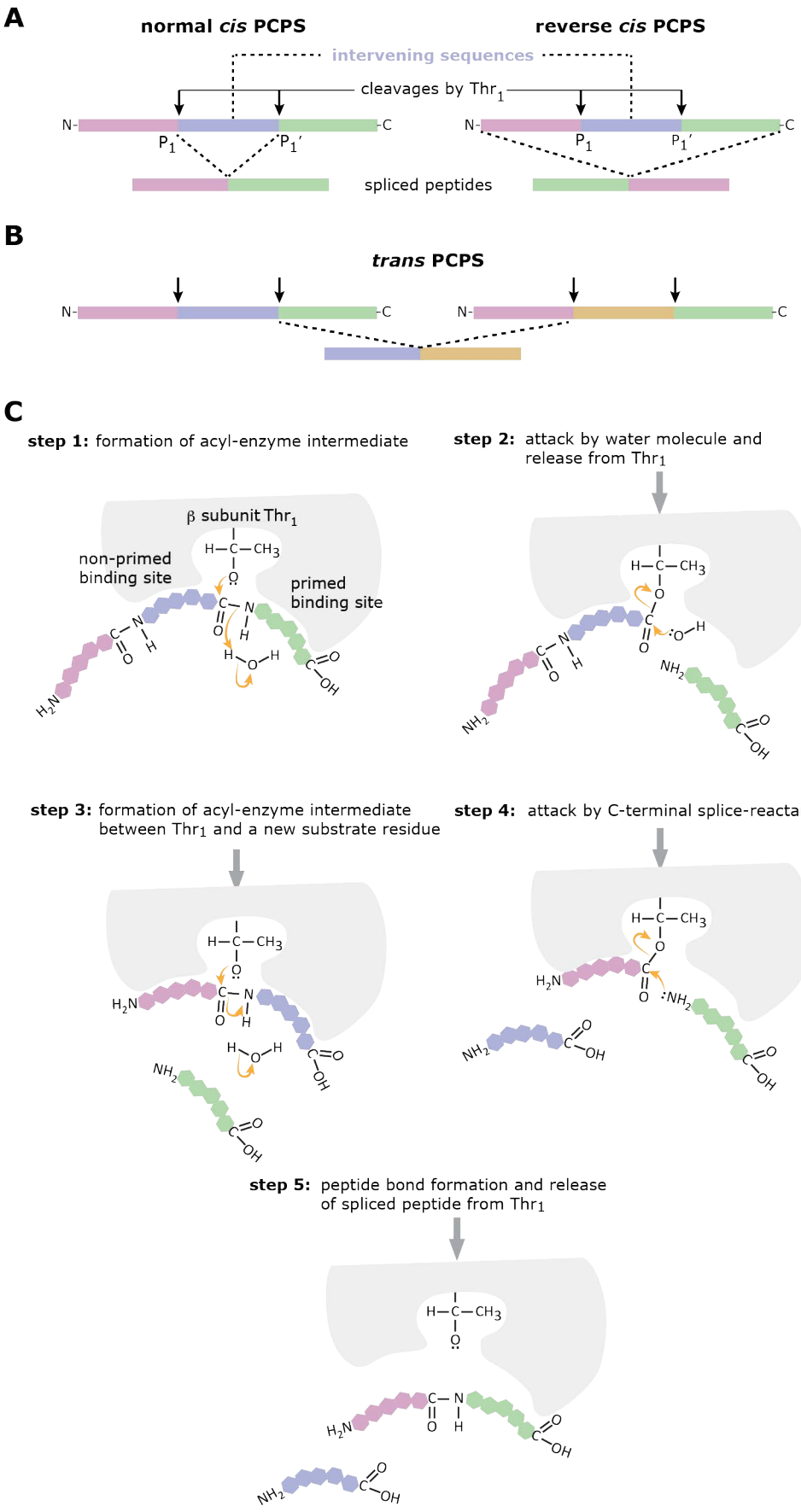
The proteasome does not only cut proteins into fragments through canonical peptide-bond hydrolysis, but also ligates them through proteasome-catalyzed peptide splicing (PCPS).

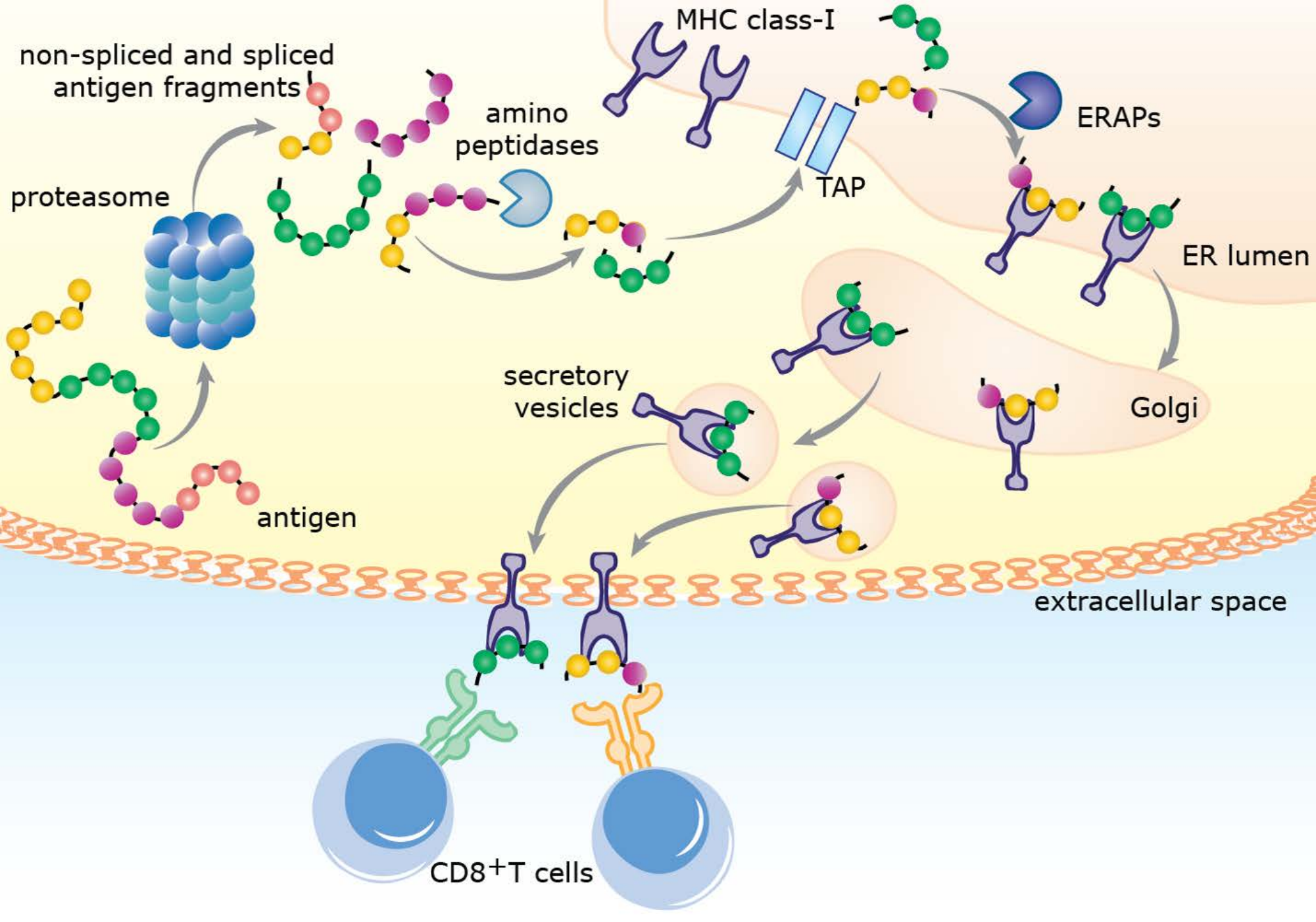
Through PCPS, the proteasome can significantly shape antigen presentation. This process is estimated to produce about one fourth of the antigenic peptide molecules and to enlarge the antigenic landscape - in term of antigenic peptide variety and antigens represented at the cell surface - by around 30%.

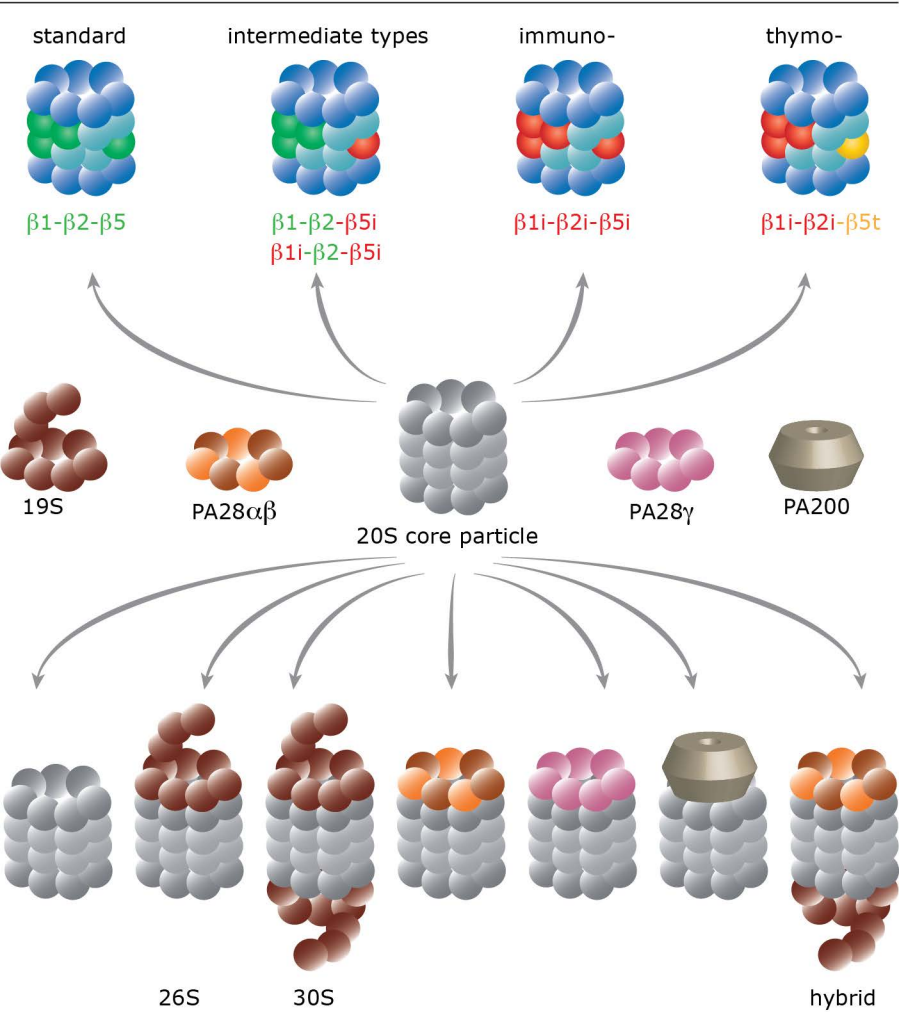
PCPS is not a random process that ligates any fragment produced by proteasome. On the contrary, it seems to be finely tuned by driving forces that have been only preliminary investigated.

Proteasome-generated spliced epitopes derived from tumor-associated antigens activate TILs, which in turn can reduce the tumor mass in patients and animal models.

MHC-I and MHC-II immunopeptidomes contain other non-canonical peptides, which can play a special role in autoimmune diseases such as T1DM


A
normal cis PCPS
reverse cis PCPS
intervening sequences
cleavages by Thr1
N-
-C
P1
P1′
spliced peptides
B
trans PCPS
C
step 1: formation of acyl-enzyme intermediate
β subunit Thr1
non-primed binding site
primed binding site
step 2: attack by water molecule and release from Thr1
step 3: formation of acyl-enzyme intermediate between Thr1 and a new substrate residue
step 4: attack by C-terminal splice-reactant
step 5: peptide bond formation and release of spliced peptide from Thr1


MHC class-I
non-spliced and spliced
antigen fragments
amino
peptidases
ERAPs
proteasome
TAP
ER lumen
secretory
vesicles
Golgi
antigen
extracellular space
CD8+T cells


standard
intermediate types
immuno-
thymo-
β1-β2-β5
β1-β2-β5i
β1i-β2-β5i
β1i-β2i-β5i
β1i-β2i-β5t
19S
PA28αβ
20S core particle
PA28γ
PA200
26S
30S
hybrid